A

Mini Project Report

On

**Data Analytics on Student Database**

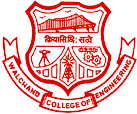
*Submitted by*

**2019BTECS00056 Rohan Bondre**

**2019BTECS00057 Shreyash Malu**

**2019BTECS00059 Sanmati Sattur**

**2019BTECS00060 Suvansh Sharma**

****

**Department of Computer Science and Engineering,**

**Walchand College of Engineering, Sangli.**

(*An Autonomous Institute*)

**2022-23**

**1. Problem Statement**

Detecting fraudulent transactions is a very significant use case in today’s scenario of digitized monetary transactions. In order to address this problem, a Synthetic Data is generated using PaySim Simulator and it is made available at Kaggle site. The data contains transaction details like transaction type, amount of transaction, customer initiating the transaction, old and new balance in Origin i.e., before and after transaction respectively and same as in Destination Account along with the target label, is fraud. So, based on the transaction details, a Classification Model can be developed that can detect fraudulent transactions.

**2. Introduction**

* Over the past years, speedy development of e-commerce techniques has been observed, making it promising for society to choose the best worthwhile product.
* This has made us dependent on financial institutions, where everyone deals with online banking facilities.
* Moreover, for payment, people are preferring online over other methods which thus, have a higher risk of getting compromised. Thus, it is a big responsibility of financial institutions to upgrade their existing mechanism to prevent these fraud actions.
* The proposed project detects fraud in monetary transaction using the XGBoost classifier to handle the imbalanced data

**3. Objectives**

Following are the objectives of the proposed dissertation work:

1. To Perform Data Analysis On Student Performance Dataset.
2. To Analyze Dataset and visualize it with pie chart
3. To find interesting patterns in dataset
4. To train model to Predict Average Score.

**4.Software Requirement**

1. Python
2. Xgboost Model
3. Streamlit
4. Synthetic Data is generated using PaySim Simulator

**5.Implementation**

1. Understanding the dataset with help of data visualization and pre-processing methods
2. Dataset contains 10,77,413 entries of transactions
3. Data Visualization
4. Insights on Dataset
5. Data Cleaning

# finding  out the types of fraudulent transaction types out of all the transaction modes.

print('The Types of fraudulent transactions are {}'.format(list(data.loc[data.isFraud == 1].type.drop\_duplicates().values)) )

# figuring out the fraud transfer and fraud cash outs

dfFraudTransfer = data.loc[(data.isFraud == 1) & (data.type == 'TRANSFER')]

dfFraudCashout = data.loc[(data.isFraud == 1) & (data.type == 'CASH\_OUT')]

1. Feature Engineering

# feature engg.

X['errorBalanceOrig'] = X.newbalanceOrig + X.amount - X.oldbalanceOrg

X['errorBalanceDest'] = X.oldbalanceDest + X.amount - X.newbalanceDest

1. Sampling

"""\*\*OVER-SAMPLING\*\*"""

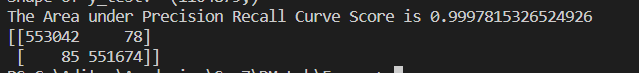
x\_resample, y\_resample = SMOTE().fit\_resample(X, Y.values.ravel())

1. Train-Test Spliting

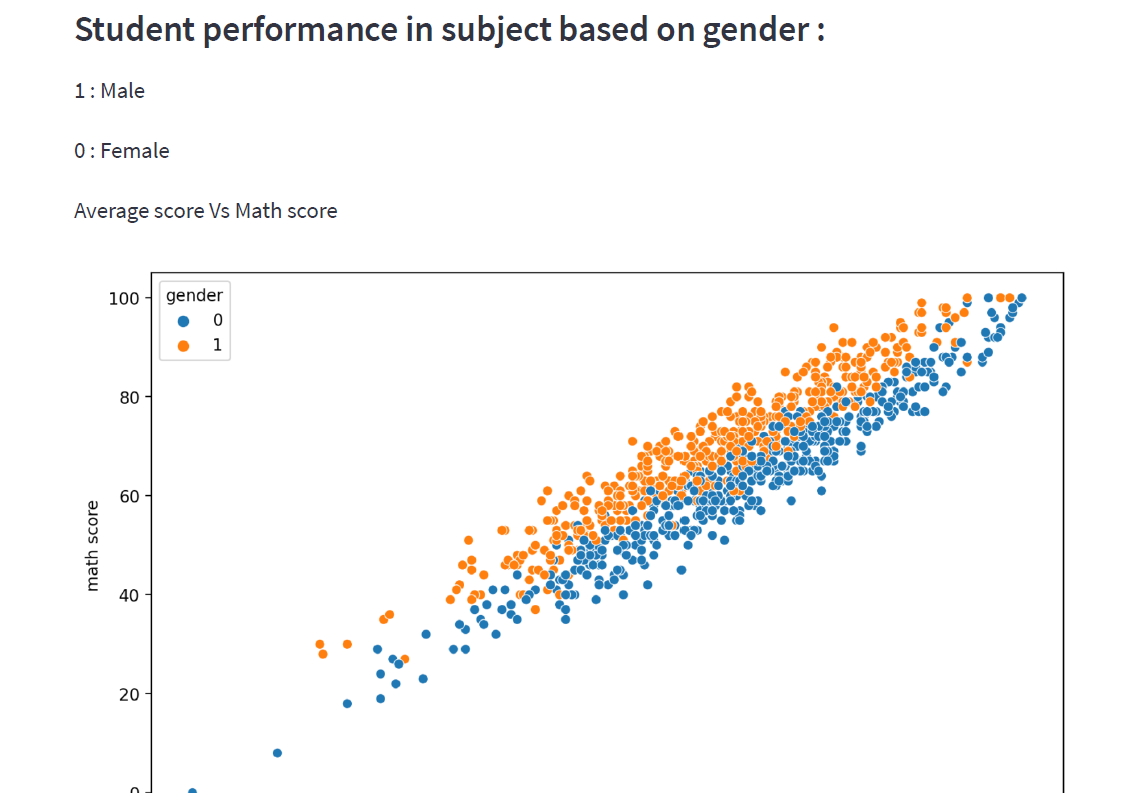
# splitting the dataset into train and tests

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x\_resample, y\_resample, test\_size = 0.2, random\_state = 0)

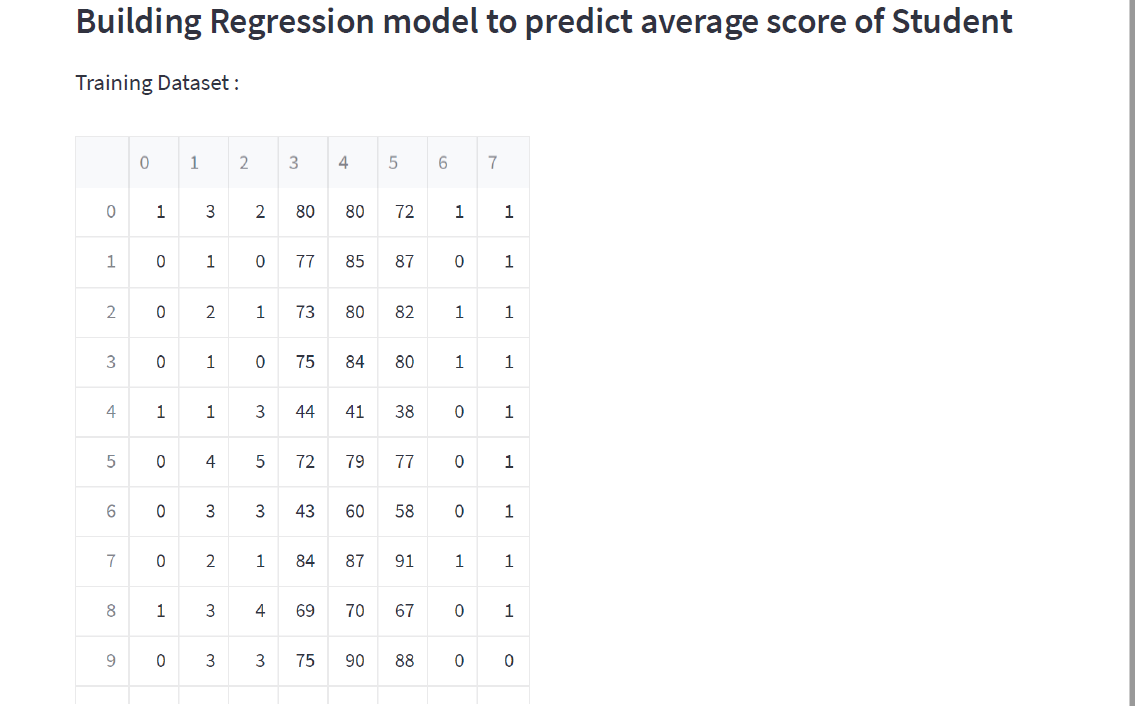
1. Accuracy – Confusion Matrix

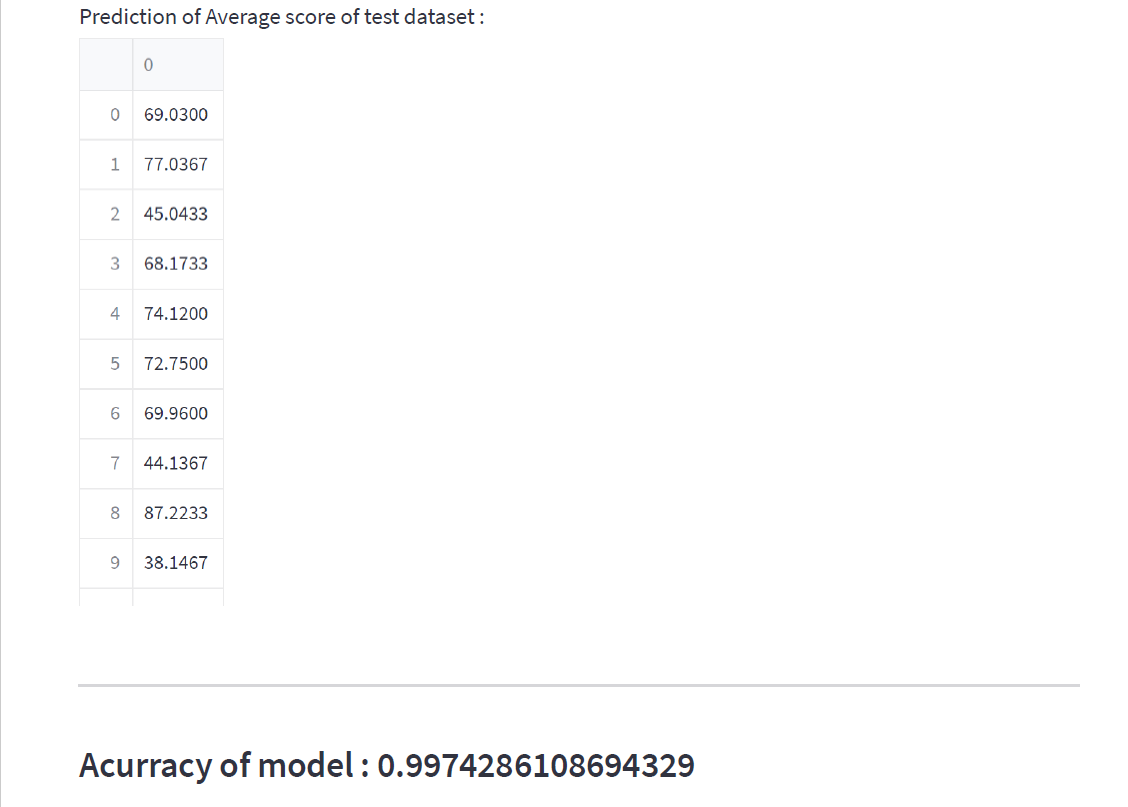


**Screenshots –**

****

****

****

****

**Conclusion**

Regression model is Built for More than 2 attributes With High Model Accuracy.

**References**

**1.** [**https://assets.researchsquare.com/files/rs-1722294/v1/e1c6dcc0-d7ca-4ffc-bebb-ae66cf310e01.pdf?c=1654889874**](https://assets.researchsquare.com/files/rs-1722294/v1/e1c6dcc0-d7ca-4ffc-bebb-ae66cf310e01.pdf?c=1654889874)

**2.** [**https://ieeexplore.ieee.org/document/9114519**](https://ieeexplore.ieee.org/document/9114519)

**3.** [**https://ieeexplore.ieee.org/document/9724422**](https://ieeexplore.ieee.org/document/9724422)